# Facade object detection using YoLoV8m

Akshay Saxena
Master of Engineering
Geodesy and Geomatics
University of New Brunswick
Akshay.saxena@unb.ca

**Abstract**

The paper discusses about the classification of façade objects such as windows, stairs, and doors within the images of buildings in street view using the YoLoV8m deep learning model provided by ultralytics library in python programming language. The GPU (Graphics Processing Unit) used is RTX 4080 mobile version. Various comparisons are made by adjusting the training parameters (image resolution, number of epochs, batch, etc.) to get the best results and efficiently detect the trained classes within input images.

## 1. Introduction

BIM is an integrating technology that provides an information backbone that transcends organizational boundaries within projects (Sacks et al., 2018). As such, BIM environments support Construction Tech innovations by mediating the gap between the information intensive technology innovations and the traditionally information poor and fragmented construction project organisations. However, BIM platforms, tools and processes are not yet able to support most technology innovations "out of the box".

Detecting building façade elements is a crucial problem in computer vision for image interpretation. In Building Information Modeling (BIM) studies, the detection of building façade elements has an important role. BIM is a tool that allows maintaining a digital representation of all aspects of building information; therefore, it will enable the storage of almost any data related to a given structure, regarding its geometric and non-geometric aspects. Façade segmentation was first studied in the 1970s using hand-crafted expertise. Later, detection and segmentation studies emerged based on shapes of objects and parametric rules. With the developing technology, deep learning approaches in object detection studies have intensified. It is obvious that the desired analyses can be performed faster with deep learning approaches. However, deep learning methods require large training data. Algorithms that consider different situations and are suitable for real-world scenarios continue to be developed. (Sezen et al., 2022).

Building façades elements detection plays a key point role in façade defects detection and street scene reconstruction tasks for sustainable city development. Although artificial intelligence technology has made a breakthrough in image segmentation, it is nontrivial to directly apply standard deep learning approaches for building façade element detection. The main reason is that the existing semantic segmentation networks have a bad performance in predicting highly regularized shapes (Zhang et al., 2022). While the Transformer architecture has become the de facto standard for natural language processing tasks, its applications to computer vision remain limited (Alexey Dosovitskiy et al., 2021).

## 2. Background

YOLO stands for You Only Look Once, and it is a single-shot object detection model that performs object detection in one pass of the neural network. Basically, there are two types of approaches for object detection using deep learning:

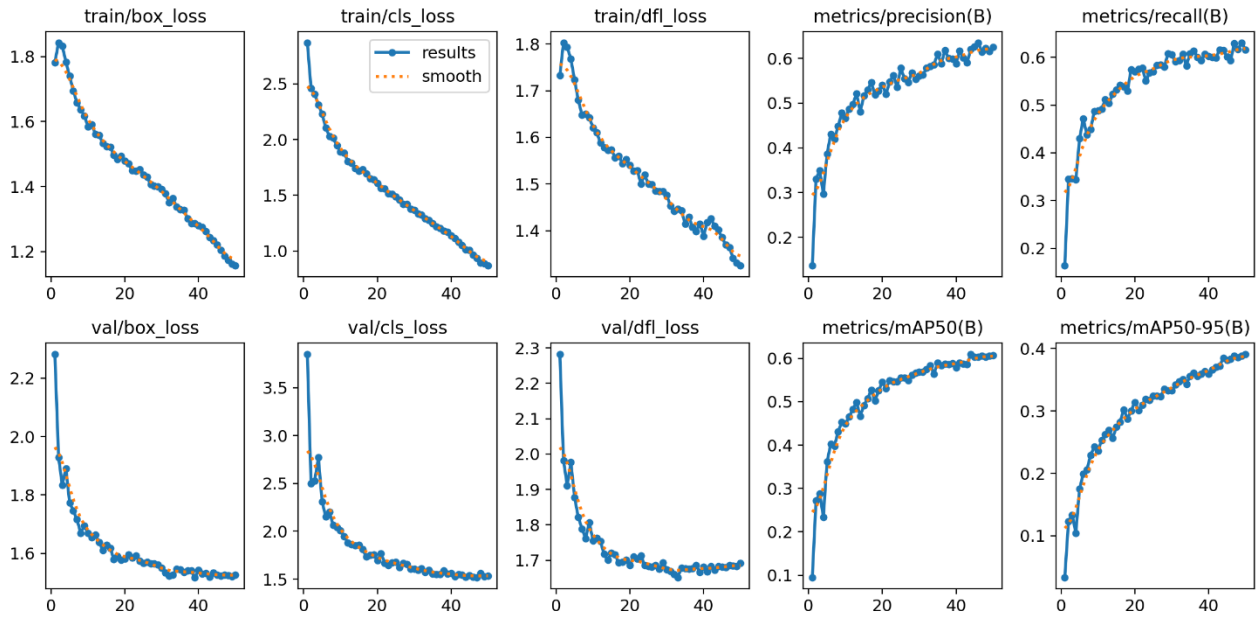
• One-Stage Detector: Models that detect, identify, and localize an object in one forward pass of the neural network which results in faster detection. Single-Stage Detector (SSD) & YOLO falls under this category.

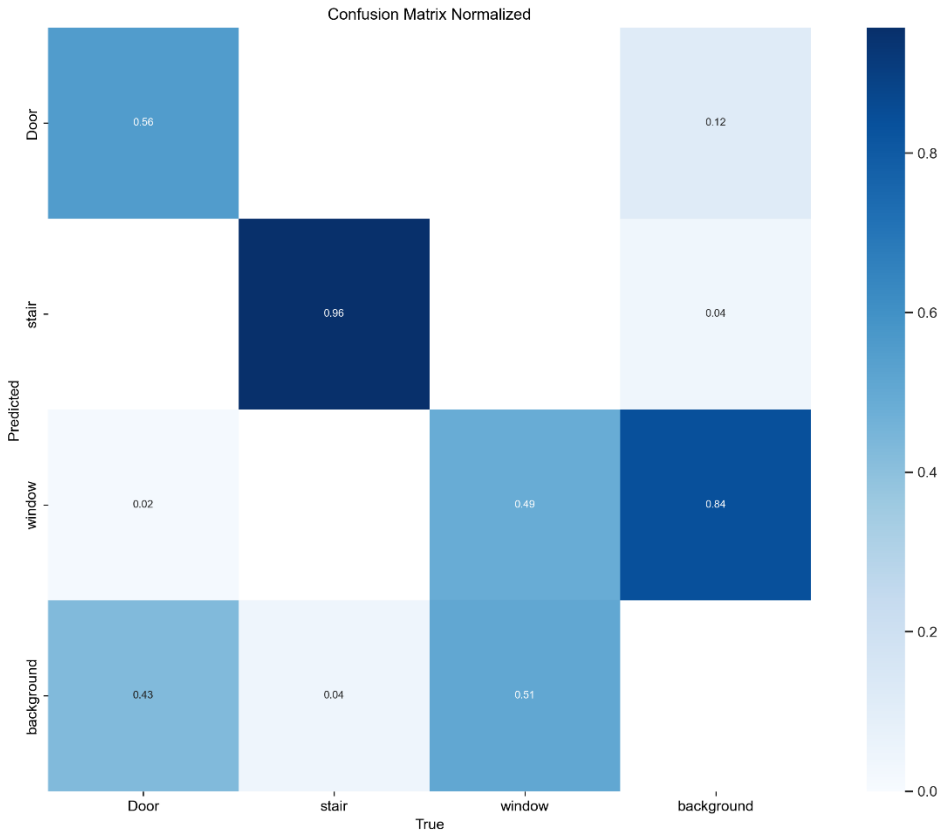
• Two-Stage Detector: Models that detect, identify, and localize an object in two passes of the neural network. In first stage, Region Proposal is done which helps model to classify the image in second stage more accurately. The two approaches are better in their own way. In Object Detection there is always a trade-off between speed and accuracy. One-stage detector has high detection speed which makes it suitable for surveillance, automatic driving, license plate detection etc. On the other hand, two-stage detector can be where accuracy is a priority like face recognition, Fingerprint, and pattern recognition, etc. (Amang et al., 2023).

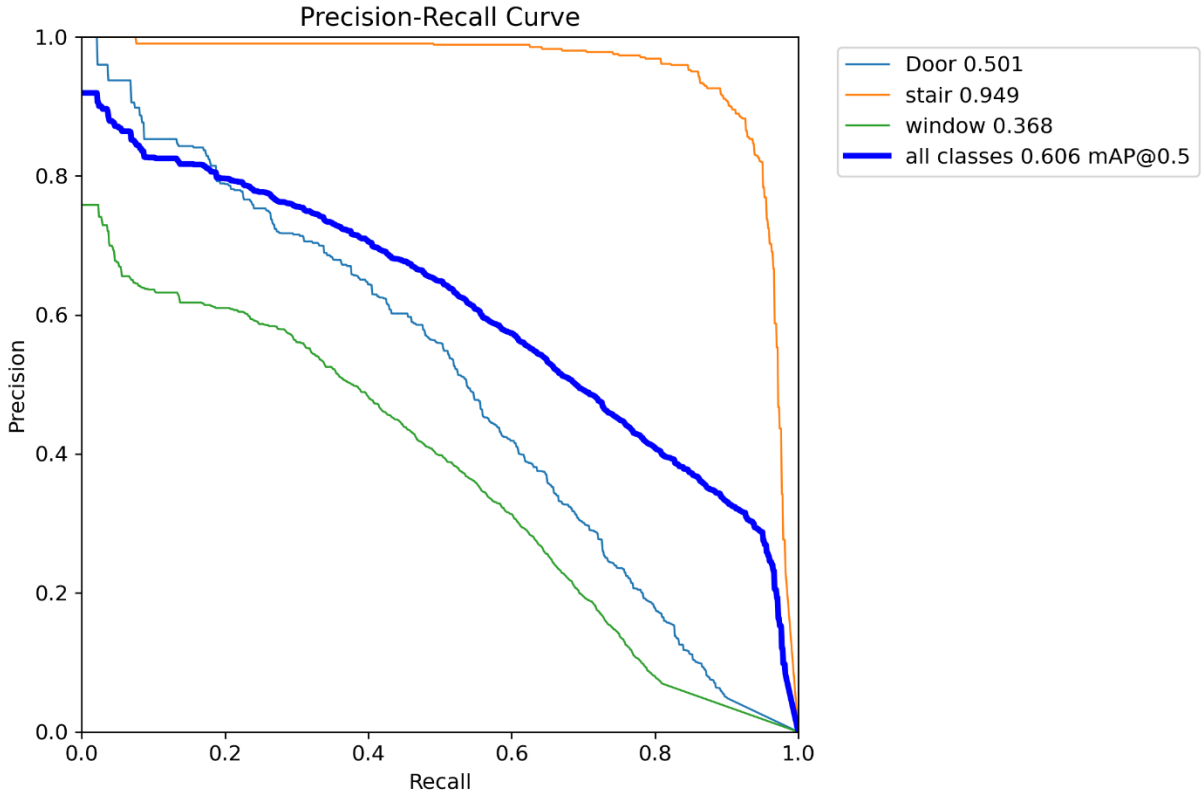
## 3. Experimental study: -

**(The number of labels per class remains same for different training parameters, doors = 1900, stairs = 1000, windows = 8000)**

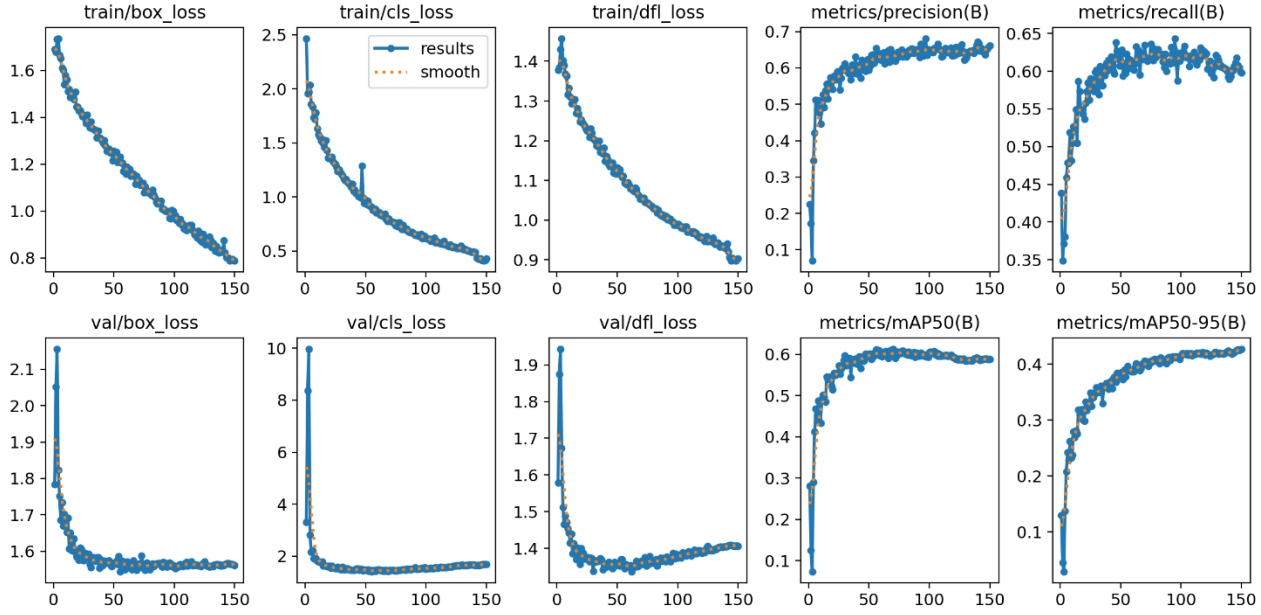
### 3.1 Image size (1280X1280), batch = 8, epochs = 50

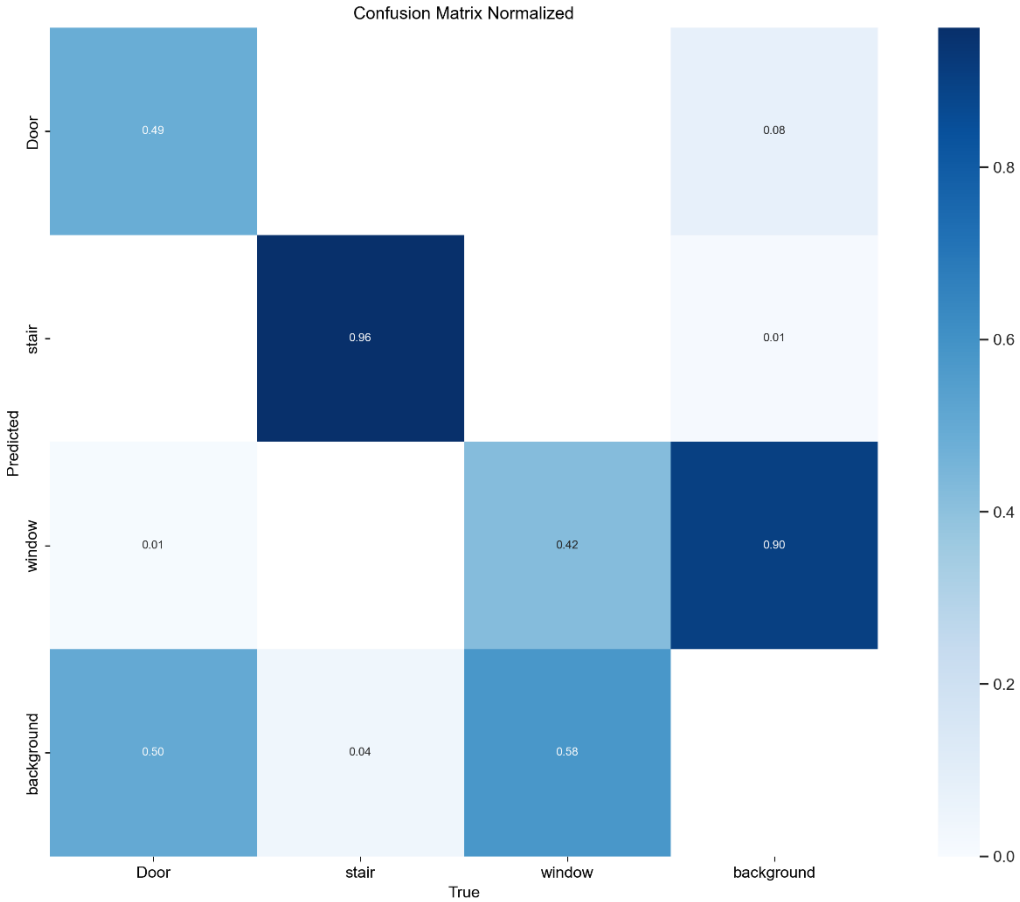


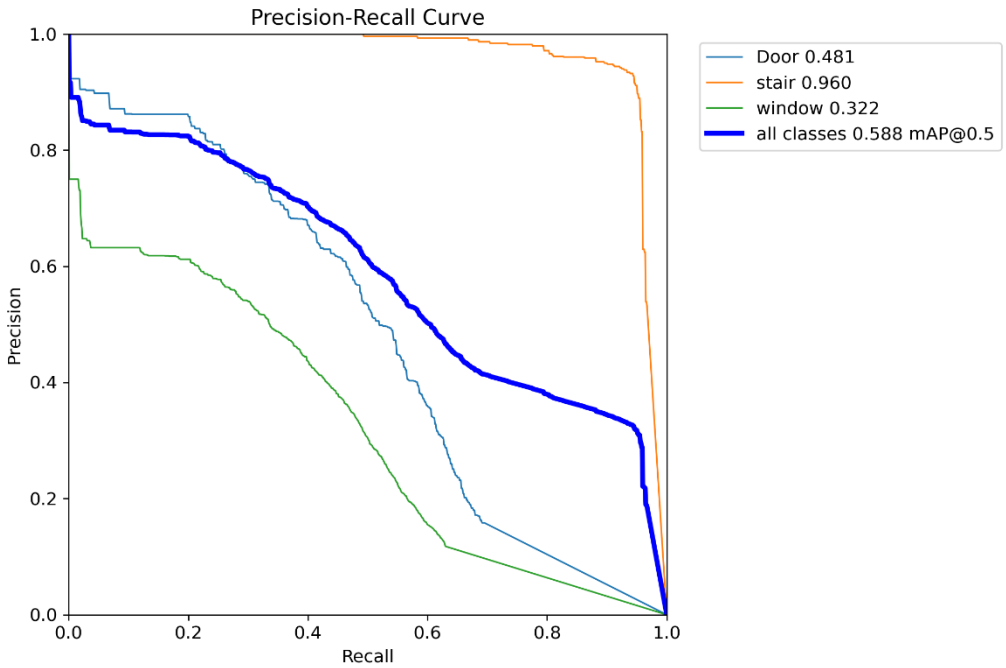

Confusion Matrix Normalized
Door
stair
window
background
Predicted
0.56
0.12
0.96
0.04
0.02
0.49
0.84
0.43
0.04
0.51
Door
stair
window
background
True
0.8
0.6
0.4
0.2
0.0


Precision-Recall Curve
Precision
Recall
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Door 0.501
stair 0.949
window 0.368
all classes 0.606 mAP@0.5

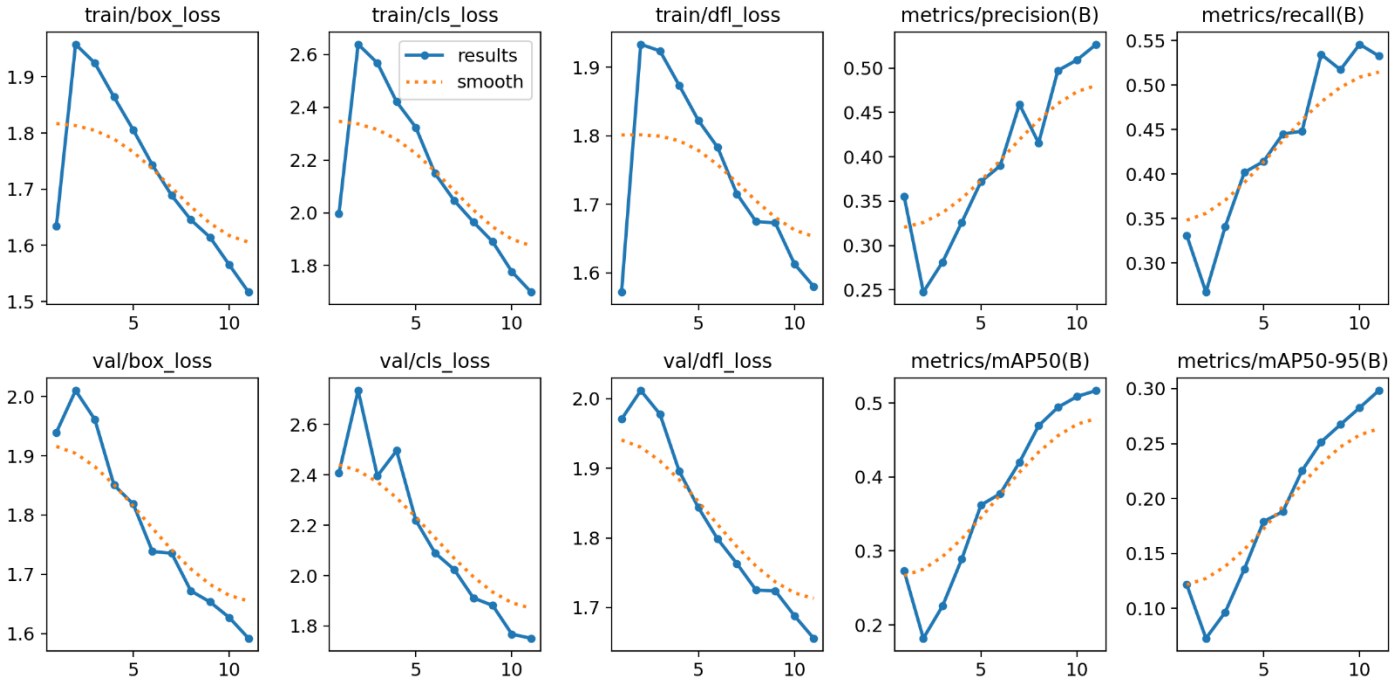
### 3.2 Image size (640X640), batch = 32, epochs = 150

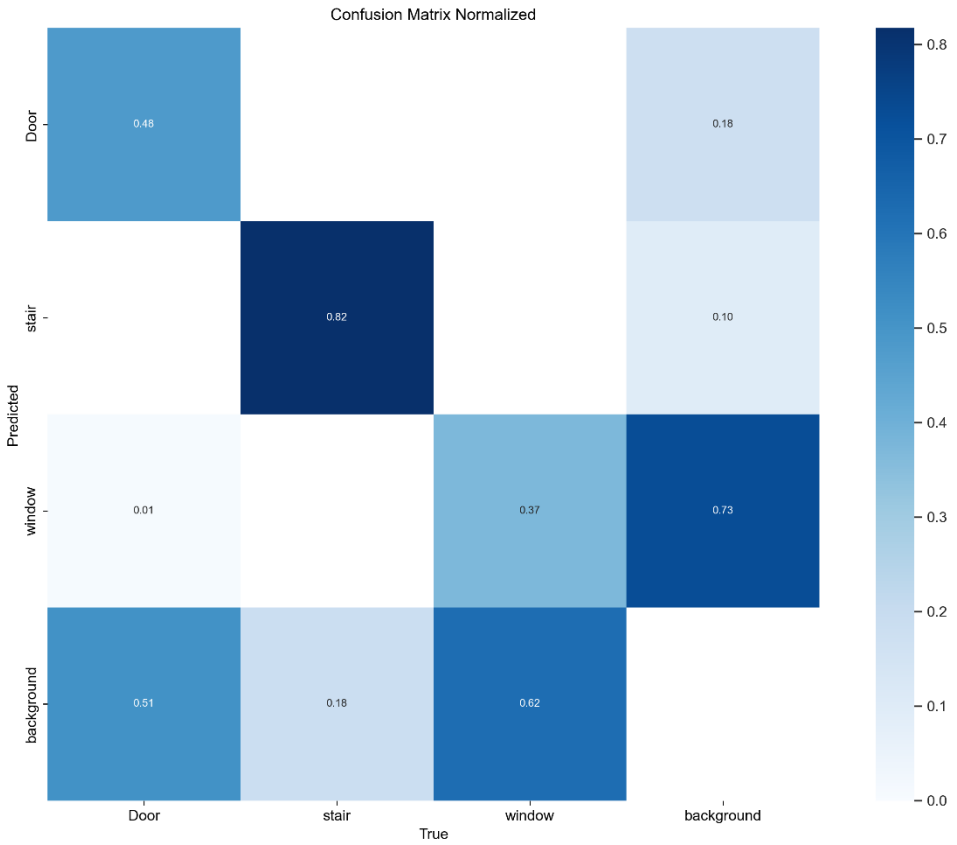


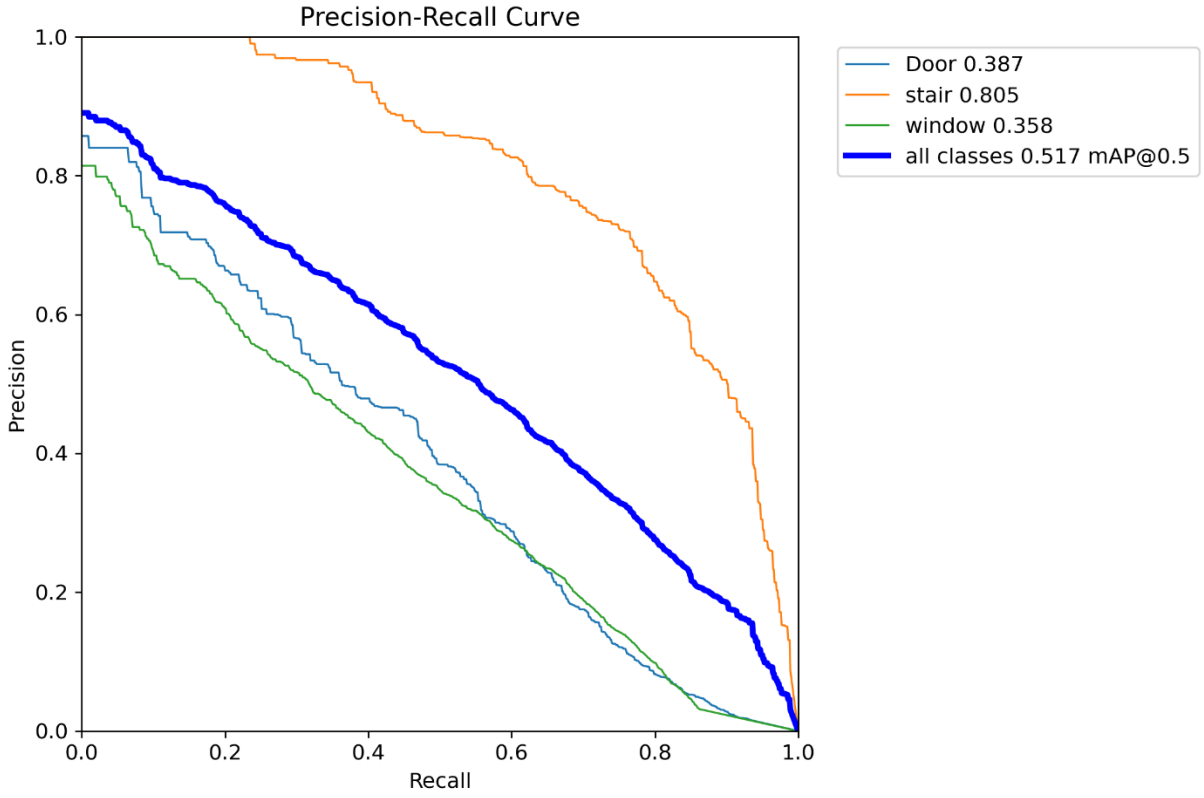






### 3.3 Image size (1280X1280), batch = 16, epochs = 10




Confusion Matrix Normalized
Predicted
True
Door
stair
window
background
0.48
0.18
0.82
0.10
0.01
0.37
0.73
0.51
0.18
0.62
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0


Precision-Recall Curve
Precision
Recall
Door 0.387
stair 0.805
window 0.358
all classes 0.517 mAP@0.5
0.0
0.2
0.4
0.6
0.8
1.0

**4. Conclusions: -**

Experiment 3.1 provides best results in classification of the objects with appropriate confidence. My future work will involve using the trained weights from YoloV8m to be transferred to Vit (Vision Transformer) using fully connected layer at the end of Vit classifier.

The YoLoV8m model will be able to classify objects such as doors, windows, stairs, etc. (as per training data contents) from the building images. This extracted information can be stored to enrich the BIM (Building Information Modelling) information on certain projects.

**References**

Zhang, G., Pan, Y. & Zhang, L. (2022). Deep learning for detecting building façade elements from images considering prior knowledge. Automation in Construction, 133, 104016-. https://dx.doi.org/10.1016/j.autcon.2021.104016

Sacks et al., 2018 R. Sacks, C.M. Eastman, G. Lee, P. Teicholz BIM Handbook: A Guide to Building Information Modeling for Owners, Designers, Engineers, Contractors and Facility Managers John Wiley and Sons, Hoboken, NJ (2018)

Sezen, G., Cakir, M., Atik, M. E., and Duran, Z.: DEEP LEARNING-BASED DOOR AND WINDOW DETECTION FROM BUILDING FAÇADE, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLIII-B4-2022, 315–320, https://doi.org/10.5194/isprs-archives-XLIII-B4-2022-315-2022, 2022.

Amang, S. ., Sen, B. ., Pradhan, A. ., Sharma, K. ., & Singh, V. K. . (2023). Enhancing COVID-19 Safety: Exploring YOLOv8 Object Detection for Accurate Face Mask Classification. International Journal of Intelligent Systems and Applications in Engineering, 11(2), 892–897. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2966